Manuscript cp-2020-79
Response to the referee 2

We would like to thank the referees for their constructive feedbacks and insightful comments. We appreciate the time and effort the referee dedicated to review our manuscript, which helped us to improve our presentation. We have incorporated the suggestions made by them, and below you find our responses to the referees' comments (in blue).

Major comments:

This is interesting work with valuable implications. However, in my opinion the authors have based their analysis and conclusions on the ability of the CESM to simulate pre-instrumental drought occurrence/frequency (and do not sufficiently prove that the CESM can do this) and draw several conclusions that appear to be based on visual comparisons of data in figures that I found hard to believe (and in some cases appeared simply incorrect) without further quantitative support.

Thanks for your comments. In the revised manuscript, we will incorporate more quantitative and constructive presentations. We will also provide more details on the ability of CESM to simulate the present and past climates, and the limitations associated with the model.

Main concerns:

1. CESM simulation data are not easily accessible without contacting the researcher who ran the simulations. (As noted below, I wanted to try replicating the authors' analysis by comparing the CESM data to the OWDA data, but the CESM data are not publicly available.)

The policy at our institute is to provide the data by request to us. We see the problem that the reviewer is thus not able to stay anonymous. For our institute infrastructure (and also for other public open data bases such as Pangaea data base), it is simply not possible to make the data publicly available due to the size of the simulation, which is roughly 100 TB. The plan though is to provide some of the data, like temperature and precipitation fields on a public accessible server, but this is not that data the reviewer needs. As the reviewer would like to compare the CESM with OWDA, we provide the self-calibrated PDSI and soil moisture anomaly for the Mediterranean region on anonymous FTP server: https://fileserver.climate.unibe.ch/public/woonkim/

2. I wonder if perhaps some parts of the Mediterranean region experience drought at different times/magnitudes in the instrumental data (also in the CESM data)? Why did the authors choose this region?
Please provide more evidence that drought/precipitation in the region varies coherently (e.g., suggest showing more information in Figure 1 other than a map and a box).

The referee 1 also commented about the same issue, thus, we provide here the same response. We agree with the referees that we need to clarify our choice of the region. We

did not mention in the manuscript that we want to study droughts with more pan-Mediterranean characteristics, as the region shows an overall drying trend in the modern period and future projection (see the figure 1 below, also some citations in the introduction of our manuscript, for example, Naumann et al., 2018). Additionally, recently some devastating drought events with pan-European characteristic, covering large area of the southern Europe, including our region of study, have been reported (Garcia-Herrera et al.,2019; Spinoni et al., 2017). Thus, understanding the mechanisms associated to these types of events in the past would be useful to understand their present dynamics and future changes.

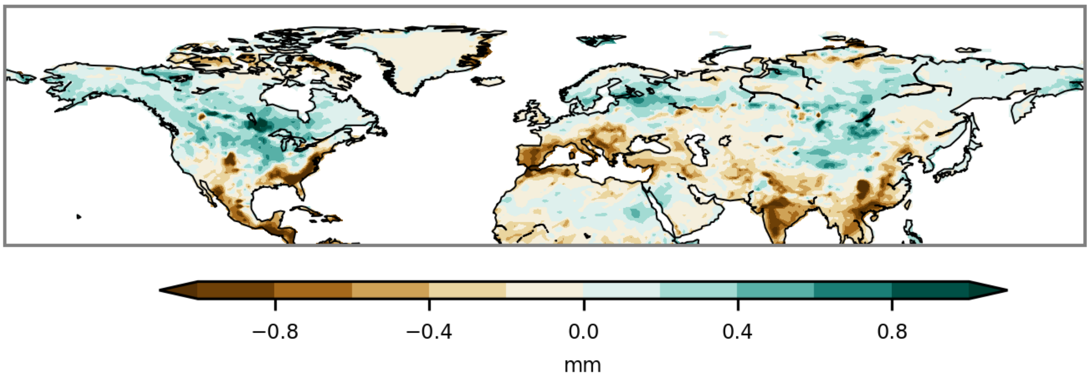

Figure 1. Mean soil moisture anomaly with respect to 1000-1849 AD for the period 1901 - 2000 AD in the CESM.

We found that when droughts occur over the region of study in the past (850 - 1849 AD) in the model, the percentage of area with the soil moisture anomaly below 0 mm over our region of study is more than 50% , in average 75.06%, covering a large part of the Mediterranean region. Thus, the droughts in our analysis can be clearly seen as pan-Mediterranean droughts (Figure 2).

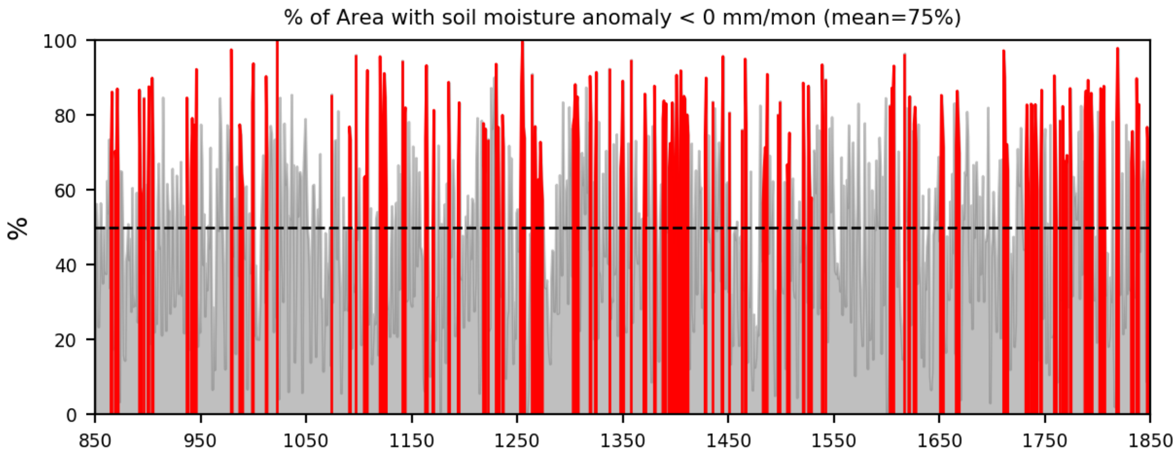

Figure 2. Percentage of area in the Mediterranean region (rectangle in the figure 2) with the soil moisture anomaly below 0 mm from 850 – 1849 AD. Drought periods are shaded in red. The mean coverage of the region with negative soil moisture anomaly during droughts is 75.06%.

In addition, the Empirical Orthogonal Function analysis on the monthly precipitation from the observation (U. Delaware V5.01 data; Willmott and Matsuura, 2001) indicates that, the chosen region shares a similar variability in the first EOF (13.28%) and second EOF (11.01%) (Figure 3). Also, this region shares the overall similar influence of NAO, which is an important driver for precipitation (Martin et al., 2004).

We will elaborate better our motivation and choice of the region as we did here in the revised manuscript.

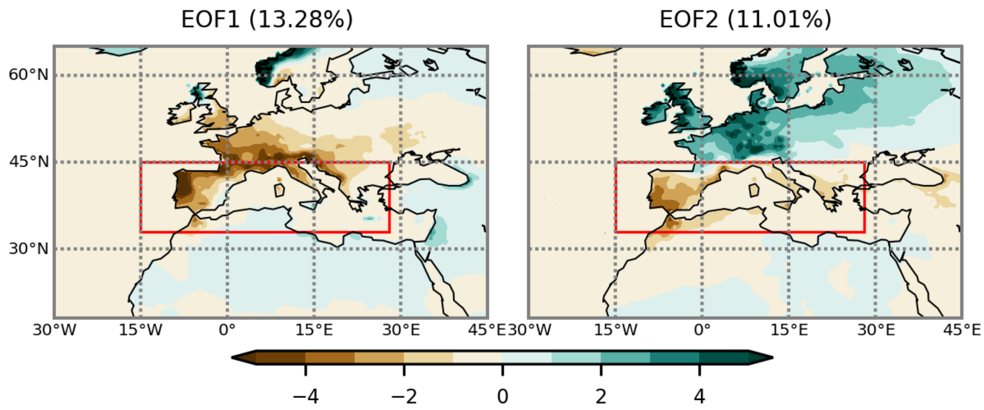

Figure 3. Variance explained by the first two EOF in the observational precipitation.

3. The authors gloss over a critical comparison of the paleo to the model data (lines 179-182) - they conclude the background drought statistics (occurrence/frequency) in the CESM are similar to the OWDA. Yet, an examination of figure 1c suggests to me that the drought occurrence in the model and paleo data are quite dissimilar- the bulk of droughts in the CESM are centered around 6-10 years in length, and in the OWDA the distribution is centered around 1-4 year drought lengths. This discrepancy is quite striking to me, and I was surprised when the authors claim these distributions are comparable.

We agree with the referee that the absolute numbers of drought occurrence are dissimilar. We will change the text according to the statistical tests we performed (See also our response #4 below). We also modified the histogram plot (Figure 2-c in the manuscript) as the figure 4 below in order to make the comparison among indices qualitatively easier.

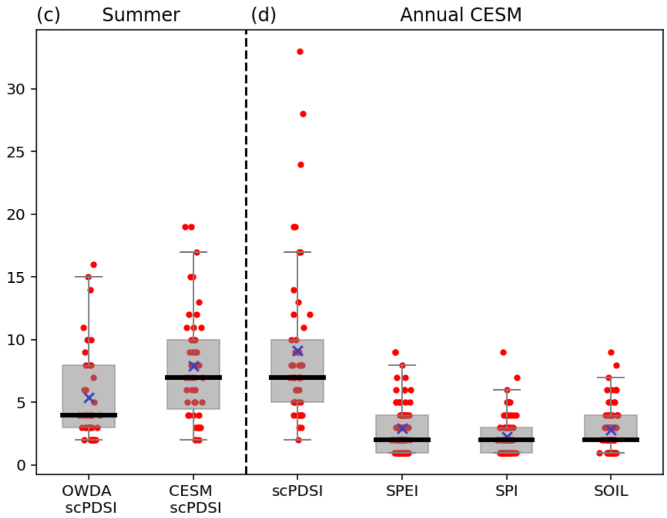

Figure 4. Distribution of durations of droughts for (c) the summer scPDSIs in OWDA and CESM, and (d) the annual indices in CESM for the period of 850-1849 AD.

4. If the authors want to make this claim, I suggest using some sort of metric (e.g., something like a Mann-Whitney or Wilcoxon rank-sum test or some sort of distribution comparison metric) to show these two drought occurrence distributions are statistically similar. Even a report of the median, mean, and range would be more helpful than the visual comparison. I also suggest the authors use other metrics such as showing average drought occurrence per century (e.g., see Figure 3 in Parsons et al., 2018, J. Clim.).

We agree with the author, thus, we performed Mann-Whitney U test and Mood's Median test to test the means and medians of indices during droughts in the last millennium. The analysis indicates that overall the means and medians of almost all indices are statistically different among each other, with few exceptions for the means between SOIL and SPEI, and the medians between SOIL and SPEI, and between SOIL and SPI. We will include this analysis in the revisited manuscript and include more discussion accordingly.

5. Other suggestions include comparing the power spectra (PSD) of the OWDA and CESM PDSI. For example, I made Southern Mediterranean regional mean time series of PDSI from the OWDA and from the CESM1 LME run2 (this is an admittedly lower resolution version of CESM1; Otto-Bliesner et al., 2015; but the background drought statistics in the CESM LME and higher resolution versions of CESM are quite similar, at least in SW North America -e.g., Parsons and Coats, 2019, JGRA) over the 850- 1849 CE time period. I found the power spectra show quite dissimilar behavior for the CESM and OWDA PDSI variables, with varying discrepancies as varying frequencies depending on how I standardize them.

We will consider this option or the inclusion of the tests in our response #4 for the revised manuscript.

6. Comparison of CESM with instrumental/reanalysis data: the authors missed an opportunity to validate the performance of the CESM in the historical/instrumental era against instrumental/reanalysis data. The authors show (e.g., Figures 3,4,6,7) background geopotential height, SST, etc. anomaly patterns associated with drought, but they have not used instrumental-based data to show the model can accurately simulate the observed climate, and I remain unconvinced the background drought statistics are similar to the OWDA (see Main Concern (3) above).

We agree with the referee that the validation of the model against the observation was missing in our study. Thus, in the revised manuscript, we will include a new section about the validation, to compare the observation-model-proxy for the period of 1901 – 2000, in terms of the numbers and duration of droughts and circulation patterns associated with droughts. Also refer to our response #7 below.

You can see in the figure 5 below, the numbers and duration of droughts in observation, model and proxy. To calculate the modern scPDSI, the gridded station data for precipitation and temperature (U. Delaware V5.01 data; Willmott and Matsuura, 2001) was used, taking the reference period of 1950-1979 AD. The same reference period was taken to calculate the 1901-2000 scPDSI in the CESM.

The model shows a mean duration of summer droughts which is relatively close to the mean from the observation (observation with 11.50 years and CESM with 9.67 years), but OWDA presents relatively shorter mean duration of 5.42 years than both observation and CESM. We did not perform statistical test here as the numbers of yearly droughts are not enough to perform a reliable test (5 droughts in OWDA, 4 droughts in the observation and 3 droughts in CESM).
We think this difference in the distributions between OWDA and CESM can explain also the difference in the duration of past droughts between the proxy and model during the last millennium (Figure 4 in our response #3). We will extend the discussion on this topic more in detail in our revisited manuscript, including the uncertainties involved in the proxy and model.

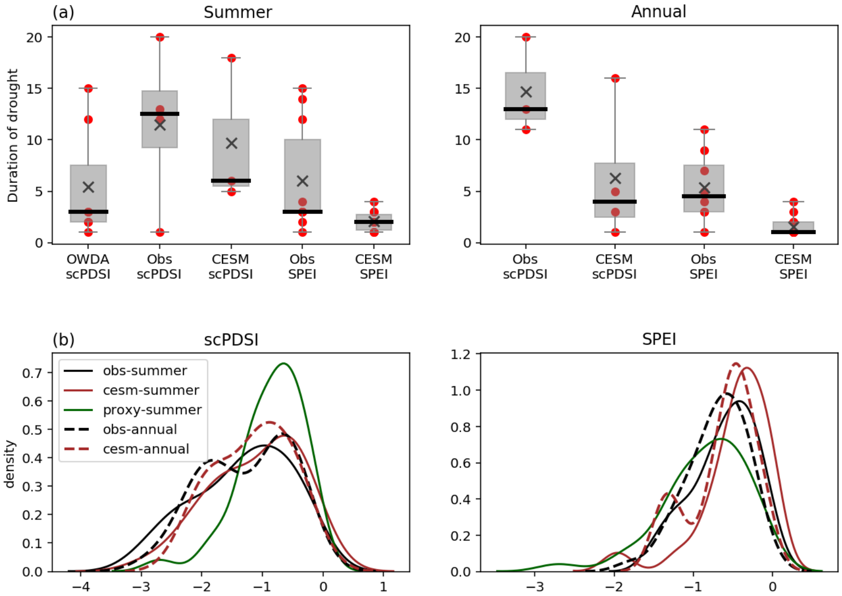

Figure 5. (a-above) Distributions of durations of droughts for (left) summer indices including OWDA and (right) annual indices. (b) Smoothed distributions by kernel density estimation of (left) scPDSI and OWDA, and (right) SPEI and OWDA.

7. Authors could compare patterns associated with drought (using a metric such as 2D pattern correlation) in the model to observed/reanalysis geopotential height (ERA5 or 20th Century Reanalysis) and SST (NOAA ERSSTv5, HadSST, etc.), as well as drought occurrence in the model to instrumental data (GPCCv2018 precipitation, Dai PDSI, CRU precipitation).Example of how other authors have made these comparisons among model and instrumental/ reanalysis data: Figure 2 in Parsons et al., 2018, J Clim., Figure 2 in Coats et al. 2013, GRL, Figure 2 in Stevenson et al., 2015, J Clim.)

We followed Parsons et al. (2018) by using the 2D pattern correlation between the scPDSI and SST (ERSST v5, Huang et al., 2017), and the scPDSI and geopotential height at 850 hPa in the CESM and reanalysis (the 21th Century Reanalysis V3, Compo et al., 2011)-observational data (U. Delaware V5.01) for the period of 1901-2000. The linear trends were removed in all variables before being correlated. You can see the result of correlation in the figure 6. Our result indicates that among the statistically significant regions, there are similar signs of correlations in the SST over the Mediterranean region and the Central Equatorial Pacific region and in geopotential height over the Mediterranean and northern Europe.

We will include this analysis in the validation section of the revisited manuscript.

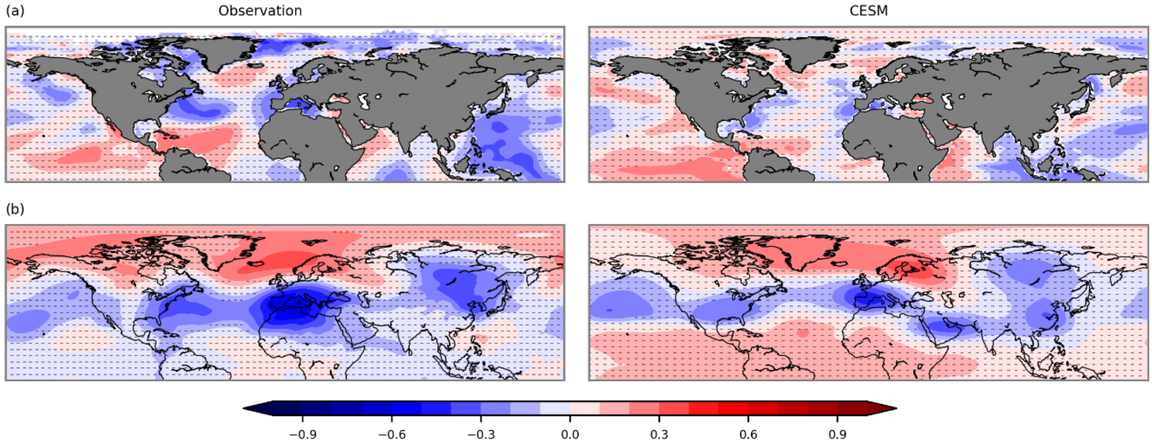

Figure 6. Fields of correlation between the monthly scPDSI and (a-above) SST and (b-below) 850 hPa for the period of 1901-2000 in (left) observational data and (right) CESM. Regions where the correlations not statistically significant at 5% level are dashed.

8. The authors do not address several of the known shortcomings in the CESM model (e.g., frequency/strength of ENSO events; Parsons et al., 2017, J. Clim, Figure 6; Bellenger et al., Clim. Dyn, 2015 for a comparison of ENSO characteristics among models and instrumental data) and what the implications of these shortcomings could be for their study, especially because the authors make claims about likelihood of ENSO events before/during/after droughts. I suggest the authors consider the findings of Ault et al. (2014, J Clim), who show that the background power spectra/statistical characteristics of drought/precipitation (e.g., white noise, power law, etc.) are critical for drought magnitude and duration, and many CMIP5-class models do not show the same background variability as instrumental/ paleo data in many regions.

We will include a discussion on shortcomings of CESM in representing ENSO and the implication of this on the drought frequency over the region in the revised version. In general, our result shows that the large circulation patterns such as ENSO and NAO, play roles in inducing droughts, though, the regional feedback may play as important role as these circulation patterns, in maintaining the persistence of droughts. We think that the latter result may not be changed by the model ENSO characteristics.

9. Especially when future warming is considered, it is important to focus on metrics of drought that don't just focus on 'atmospheric centric' supply and demand, especially if ecosystem/water resource drought impacts are important. See Swann et al., 2016, PNAS, and Swann (2018) who note that drought severity/impacts in a warming climate can be grossly overestimated by use of variables/metrics such as PDSI.

We agree with the referee that drought indices that are based on the atmospheric supply and demand, mostly associated with the temperature, have problems of overestimating future dryness, which is the case for the scPDSI and also the SPEI (Mukherjee et al., 2018). We cannot avoid using these indices, as they are the most commonly used indices and calculable using the model variables. Many of the indices proved to overcome the warming climate problem (such as indices based on vegetations), are not easily deductible from the

model variables we have. This is also the reason why we chose the soil moisture anomaly to describe droughts after the index comparison in our study (section 3.1 in our manuscript). We will include more detailed discussion on this problem in the result and discussion section of the revisited manuscript.

10. I appreciate that the authors included 10cm soil moisture, but given that surface soil water content can basically just follow precipitation variability in many regions, and thus, not really reflect full depth soil moisture trends (e.g., Berg et al., 2016), I think it would be helpful for the authors to show that they are analyzing variables actually relevant for plants/ecosystems/water resources in a warming climate, and not just supply/ demand from the atmosphere. At least a discussion of some of these points could really strengthen the paper.

Thanks for your suggestion. We will elucidate the relationship among vegetation, soil and hydrological dynamics, their changes and problem in the warming climate in the results and discussion section of the revisited version.

Specific comments:

11. Lines 13-14: the authors just list one or two types of drought (meteorological), but what about hydrological, agricultural/ecosystem, socioeconomic types of drought?

Here, we mostly used the drought metrics that reflect the meteorological (SPI), and agricultural (scPSDI, SPEI, SOIL) droughts. Though, it is important to mention that when the time scales of droughts become longer (like 1 year as our study), the differentiation among types of droughts becomes more difficult. We will include more explanation on different types of droughts in the introduction of the revised manuscript.

12. Line 22: 'climate hot-spot'- please cite a paper that shows this
Line 23: 'increase in drought episodes' – again, please cite a paper supporting this
Lines 45-46: 'attributed to the increase in the atmospheric greenhouse gases (GHG) concentration, which causes . . . decrease in precipitation over the region' - citation?
Line 52: 'unprecedent intense drought projections' – citation?

We will correct the respective citation in the revisited manuscript.

13. Lines 63-64: The separation of ocean-atmosphere conditions during various drought stages has been done before- nice to acknowledge previous work (e.g., Parsons and Coats, 2019; Namias, 1960).

We will update the citations in the revisited manuscript.

14. Lines 76-77: 'warm-dry temperature-hydroclimate co-variability at multidecadal timescales' confusing wording.

We will remove the "warm-dry" part and reformulate the sentence.

15. Line 92: 'high horizontal resolution' is subjective (and now closer to 'average resolution') in many CMIP6 models).

We will remove "high" and replace it by the resolution of the model (1° x 1°).

16. Line 102: Why not use the CESM LME (Otto-Bliesner et al., 2015)? There are more iterations, with several RCP8.5 extensions (and a much longer 1000 yr piControl run that is easier to compare w the last millennium runs given the similar length of simulations), allowing for a more complete analysis of internal variability. Is the background climate state that much better in the 1 degree vs the 2 degree version of the model? I ask because the authors explicitly state on lines 119-120 that they are interested in studying internal vs externally forced variability, and multi-model ensembles provide an ideal experimental framework for doing this.

We agree with the referee that we should provide more explanation on why we used this single model and no other available model output.

First, the spatial resolution of this long simulation is a great advantage for our study in a small constrained area. The number of grid cells over land is 256, and this would decrease roughly four times if we used a simulation with 2 degree resolution. Though we used the regionally averaged indices, we think that the spatial resolution allows a better representation of regional processes relevant for precipitation, therefore, for droughts. For example, the rainfall over the region is strongly influenced by extratropical storm tracks and cyclones. The precipitation and atmospheric dynamics associated with these climatic features depend on the model spatial resolution and better represented in GCMs with higher resolutions (Champion et al., 2011; Watterson, 2006). Hence, we think using a model with finer resolution is appropriate for this region.

Second, we want to study the physical mechanisms of continuous Mediterranean droughts during the last millennium. One of our main focus is to study the roles of the internal variability vs externally forced variability, and we believe that we were able to answer partially to this question using this CESM simulation. However, we are aware that more analysis is needed to support this result, for example, to assess more clearly the role of volcanic forcing on droughts. For that, we plan to apply some more methods, such as a wavelet coherence analysis, in a similar way as Coats et al. (2013), between drought indices and volcanic eruptions.

We will include the new analysis and an explanation on the benefit of using this model. To be balanced, we will also discuss potential drawbacks and limitations of a single model study.

17. Line 107: the years 2001-2020 AD/CE are not the future

We will modify the term to "present".

18. Lines 103-112: Suggest just citing Lehner et al. for the model description

The referee 1 commented that we should include some details on the model, specifically on the soil component. Thus, we will keep this part as it is now to provide a brief information to the readers on the model.

19. Line 127: As in Main Concern (2), please show the region varies coherently in instrumental/paleo and the version of CESM used here

Refer to our response #6.

20. Lines 131-132: removing a linear trend over the 1850-2099 time period looks quite problematic to me (e.g., Figure 9)- removing a linear trend over this time period will add in non-climatic variability artifacts from the trend removal. It looks to my eye like there is a trend 1900-2000, then a separate trend 2000-2099.

We applied the detrending method to two time periods separately: the 1850-2000 and the 2001-2099. We mentioned this procedure in the Analysis and Methods section (2.2) in lines 132-134. It is true that the detrending method can insert some non-climatic artifacts in the processed time series, but we think the way we performed the method can alleviate this problem. The figure 9 in the manuscript shows the precipitation and soil moisture anomalies with respect to 1000-1849 AD before being detrended. We will make this clearer in the revised version.

21. Line 149-150: linear temperature trend is removed, but then authors study the impacts of warming using this drought metric, which includes temperature. . .so have the authors removed temperature changes, then try to study the impacts of warming on drought? This reasoning doesn't make sense to me. Perhaps a more clear explanation. of trend removal would help (?).

We removed the trends in variables, in order to see the secondary drivers of droughts in the future excluding the effects of warming, and whether these secondary circulation patterns associated with droughts in the future are still similar to those during the last millennium. We were not clear at delivering this point, thus, we will elaborate more about it in detail in the revised manuscript.

22. Lines 140-155: As in Main Concern (6): I think all of these drought metrics/variables, with the exception of upper 10cm soil moisture, do NOT reflect actual impact on plants/ecosystems in a warming climate. Also, upper soil water content can diverge from deeper soil water – authors should show that this is a useful metric here that is distinct from precipitation alone if they want to argue that their study has relevance for ecosystem impacts.

We will include more discussion on the complexity involved in drought metrics in our new manuscript. Also refer to our response #9.

23. Lines 161-164: This drought counting method appears similar to Herweijer et al. 2007; Coats et al. 2013b- did the authors come up with this metric, or can they use a similar metric to previously published work (if so, please cite) to maintain consistency across the literature?

The drought counting method we used differs from the counting methods from the literature mentioned here including the one by Herweijer et al (2007). In our work, one drought cluster has to have only negative or zero anomalies with at least one year that the drought index falls below its 10 percentiles of distribution, and any wet year would stop the continuity of drought. By defining a drought cluster in this way, we make sure we only take strong events, also, assuring that a dry condition persists throughout the entire year. We explained our counting method in the 2.3 section of our manuscript.
We have already checked that the other counting methods (for example, the method by Coats et al. (2013) that a drought starts with two continuous years of negative anomalies and stops with two continuous positive anomalies) are not appropriate for our region of study, where slight dry conditions are more frequent.

24. Lines 168-170: see above note about similar methods in Parsons and Coats as well as Namias.

We will update the citations and texts accordingly in the revised manuscript.

25. Lines 179-183: As in Main Concern (3): Please be more quantitative. To my eye, these distributions do not appear similar- the OWDA shows droughts that are mostly 1-4 yrs, and the CESM shows droughts centered around 8 yrs. Please use a more quantitative method to compare drought time series power spectrum and/or drought frequency in paleo and model data.

We modified the plots and performed some statistical tests. The result will be updated in our revised manuscript. Refer to our response #3 and #4.

26. Lines 187-188: difficult to visually compare these different drought metrics in lower panels in Figure 2 because the x axis limits are different.

We modified the respective plot, and this will be incorporated in the revisited manuscript. Refer to our response #3.

27. Lines 204-205: 'no noticeable changes in occurrence of droughts' - is this to the eye? Can you use a more quantitative method to show this (e.g., running counts of droughts in 50 yr windows or something like that)?

We agree with the referee that the text is misleading. We wanted to say that "no noticeable coherent changes among indices during different climatic periods, such as the MCA and LIA.". However, we are aware that we should provide more quantitative analysis to support this statement. We will include the plot on running mean of duration (which you can already see in the figure 7 below) and perform a wavelet coherence analysis to address more clearly the role of volcanic eruptions on droughts. Also refer to the third paragraph of our response #16.

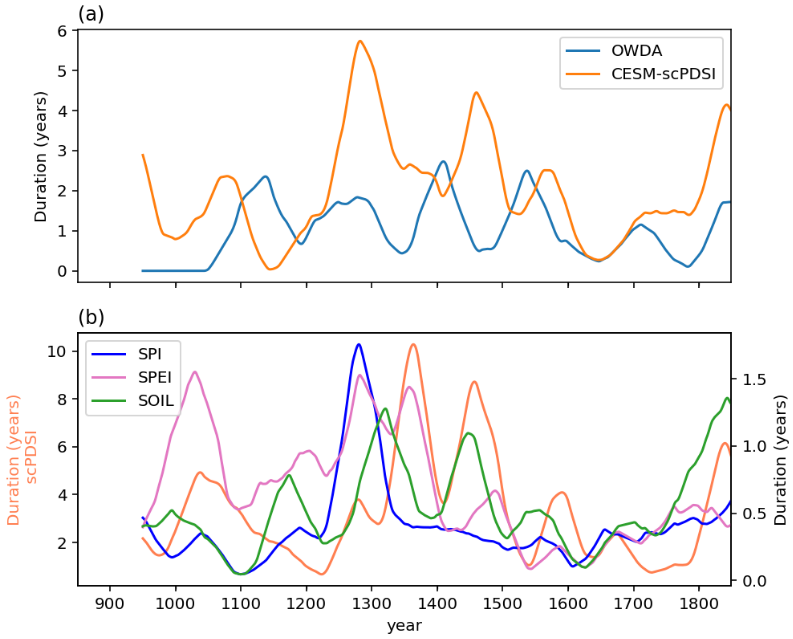

Figure 7. 100-year running mean of duration of droughts given by the (a) summer indices: scPDSI from CESM and OWDA, and (b) annual indices from CESM. Note that in (b), the scPDSI has different duration-scale (left y-axis) to other indices.

28. Lines 205-206: 'not driven by external forcing': again, this conclusion appears to be drawn based on a visual comparison, which seems insufficient to me. Lehner et al. (2015, ESD, Figure 5) use running correlation to compare model output, which I imagine could be applied here, as could some sort of wavelet/coherence analysis between volcanic forcing time series and the OWDA and CESM data. Also, Superposed Epoch Analysis or Composite Analysis could be used with volcanic forcing time series/large eruptions. At minimum, it would be great to see a time series showing the external forcing to be able to compare to the drought time series in Figure 2.

We plan to perform a wavelet analysis and address more clearly the role of volcanic forcing on droughts. Also refer to the third paragraph of our response #16.

29. Line 209-210: sentence wording is confusing/complicated.

We modified the corresponding sentence as: "Next, we investigate the underlying dynamics which are associated with Mediterranean droughts. This analysis uses the SOIL as the drought indicator.".

30. Lines 211-215: So if the r value is 0.78, doesn't this imply that only 60% of variance is shared by the two time series?

Yes, in which we think it is a good percentage of variance shared by two time series.

31. Lines 218-220: 'control simulation presents 29 droughts'- this comparison with the transient simulation is non-sensical/misleading given the two simulation lengths are different. Can the authors instead present the average numbers of droughts of various lengths per century (e.g., Parsons et al., 2018; Coats et al., 2015, Figure 5). This gets around the issue of having different length time series and gives more meaningful information about drought risk standardized to a given time window (e.g., number of droughts per 100 or 500 years).

We will change these numbers by mentioning the number of events/100 years.

32. 222-224: Is this the first time these patterns have been presented? Seems that a paper like Markonis et al. 2018 (Nature Communications) or other similar papers have previously presented similar patterns associated with hydroclimatic variability.

We mentioned in our manuscript that a similar pattern is found in Xoplaki et al.(2003) (lines 239 - 254).

33. Lines 229-236: Similar to the point I raise in Main Concern (5)- It is well documented that this model simulates ENSO events that are too strong and too frequent (e.g., Bellenger et al., Clim. Dyn., among others)- how does that impact these results? For example, if the model simulates too strong, too regular ENSO events that unrealistically influence global climate, then is it surprising that a signal from ENSO is apparent in European drought/climate? And is this finding meaningful if it's based on model bias?

Refer to our response #8. We will include more discussion on the ENSO related issues in the model and the possible impacts of them in the revised version. The response of ENSO in European climate has been addressed in some literature before (Mariotti et al., 2002; 2008; Brönnimann, 2007; Brönnimann et al., 2007), which we mentioned in the introduction of our manuscript.

34. Figure 3 caption: the caption states 'means are not statistically significant'- unclear. Please be more specific. Also please clarify if data are annual, JJA, etc. in figure caption. Additionally, the significance dots are nearly impossible to see on the dark red/blue background

We will update the captions and modify the respective plots in the revisited manuscript.

35. Lines 246-250: Are these % changes in drought/rainfall meaningful (e.g., for agriculture, ecosystems), or do these changes fall well within normal climate variations that don't have a large impact?

In the future projections for 2070-2099 AD in CMIP3 model, the Mediterranean region shows a strong drying trend with the decreases in mean precipitation of 15.5% (for the wet season with a decrease of 9.7% and dry season with a decrease of 23.6%) relative to 1950-2000 AD, and this change in water cycle would have strong impacts on the region – on society, ecosystem, etc (Mariotti et al., 2008). In our study, the rates of decrease in precipitation during droughts are 13% (winter-spring) and 11% (summer-autumn) with respect to the mean non-drought periods, and these values are in range with the rates

presented by Mariotti et al. (2008). We will add a discussion on these values in the conclusion of the revised manuscript.

36. Also, is the background variability (e.g., standard deviation, mean) of rainfall in the CESM realistic, or can we chalk this up to model bias?

We compared the summer and winter mean precipitation fields, and the mean annual cycles over the western and central Mediterranean region between the observational and CESM data for the period of 1901-2000. We applied the Mann-Whitney U test to see the significant grids at 5% of confidence level. In general, the observation and model show a similar pattern, except for few grids in the summer and different grids in the winter (Figure 8). For the annual cycles, it seems that the model shows overall less intense precipitation than the observation, but the maximum and minimum months of precipitation are coherent (Figure 9). Thus, we think CESM is good enough at performing the present rainfall over the region. We will include this comparison in the validation section of the revised manuscript. Also refer to our response #7.

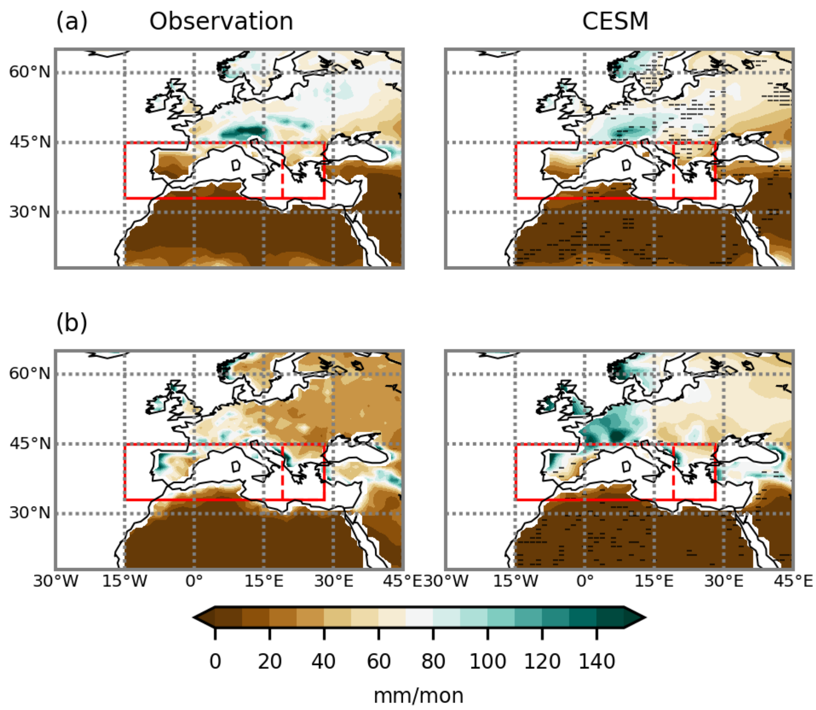

Figure 8. Mean seasonal precipitation during the (a-above) summer and (b-below) winter for the (left) observational data and (right) CESM. In CESM, Regions where the means are not statistically significant compared to the observation are dashed.

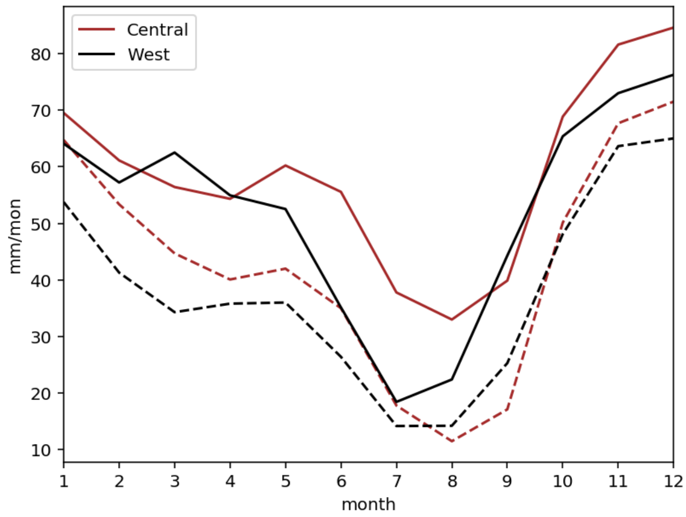

Figure 9. Mean annual cycles of precipitation over the regions in figure 8, in the observation (continuous line) and CESM (dashed).

37. Lines 254-255: similar to Main Concern (4), what about in 20th century reanalysis, ERA5, or some similar reanalysis product vs GPCCv2018 or CRU precip? Or Dai PDSI?

Refer to our responses #6 and #7.

38. Lines 257-260: 'The starting point is. . .to one or both of them'- confusing wording

We will change the respective paragraph in the revised manuscript.

39. Lines 262-269: So in other words, there is about equal odds of being in a drought during various NAO or ENSO phases? This seems important because the authors claim on lines 294-295 that a certain combination of NAO and ENSO conditions are important for initiating drought. . .but it appears to me as though there are nearly equal odds of this happening (60%) based on the phase of NAO/ENSO. Is this interpretation incorrect?

We will update this part of analysis, also due to the comment from the referee 1. We modified the thresholds to discern negative and positive phase of ENSO and NAO. Instead of using the absolute values of -0.5 and 0.5 for ENSO, and negative and positive anomalies for NAO, we set the thresholds based on the percentiles of the distribution during the non-drought periods: 25 percentile for negative phase and 75 percentile for positive phase, both for ENSO and NAO. You can see the modified plot in the figure 10 below. The preference toward positive NAO and negative ENSO during droughts is more clearly observed with this way of classification. We will change the text in the result section according to this new classification.

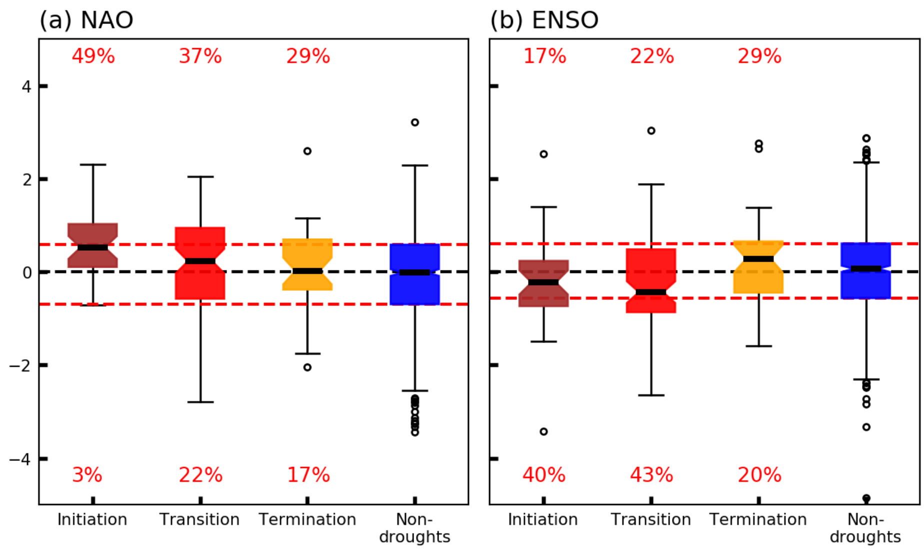

Figure 10. Distribution of (left) NAO and (right) ENSO during the initiation, transition and termination stages of multi-year Mediterranean droughts and during the non-drought period. Dashed red lines indicate the 25 and 75 percentiles of the distribution during the non-drought periods, and the numbers on the plot show the percentages of events that are under or over these thresholds in each stage of droughts.

40. Lines 298-310: I don't see how Fig 8 proves the point. Basically, it looks to me as though drought starts off dry and then transitions to less dry conditions at end of drought, and this is distinct from wet years.

We agree that droughts start off dry and become less dry with time. Here, we want to emphasize the importance of development of anticyclonic center associated with the high geopotential height anomaly which would be the driver that maintains the dry condition over the region for long time while the roles of the large-scale circulation patterns are decreasing from the transition to the last stage of droughts.

41. Lines 325-327: Similar to Main Concern (4); I have not been shown how the model performs compared to instrumental/reanalysis for the relevant variables over Europe/Mediterranean, so these conclusions don't mean a lot to me.

Refer to our responses #6 and #7.

42. Lines 337-340: 1) I see no major changes in distribution of drought in Figure 10- are these distributions distinct? Please see previous comments related to statistically distinguishing distributions (and not visually distinguishing), especially when they appear to overlap. 2) Any future changes in ENSO in this model should be interpreted with caution as most CMIP5 models, including this one as far as I can remember, struggle to reproduce the observed trends in the tropical Pacific (see Coats and Karnauskas, 2017, GRL as well as Seager et al., 2019, Nature Climate Change).

Regarding the point 1), we performed the statistical tests between the means of detrended variables for the future and for the last millennium period, and we found they are

statistically indifferent. We will include some sentences mentioning about this result. Regarding the point 2), as the referee says, we will be more careful at interpreting ENSO in the future scenario. Still, our analysis shows that the ENSO during droughts does not exhibit significant statistical changes in the future compared to the last millennium when the linear trend caused by warming is removed. This may indicate that the background ENSO during droughts are still the same. We will include more discussion on the ENSO-related issues in GCMs in the revised manuscript. Also refer to our response #8.

43. Lines 344-345: As figure 9 shows, trends in the region are not linear 1850-2100, so trend removal is problematic over this time period. Perhaps it makes sense to remove the trend 2000-2099, but otherwise the authors could be adding an artifact of trend removal into the analysis.

Refer to our response #20.

44. Lines 358-359: 'our analysis shows that the overall similarities': as stated above, the authors never actually showed this statistically, just a visual comparison.

We will update the text according to the added analysis in the revised manuscript. Refer to our response #3.

45. Line 383: 'climate over the region to a drier climate have started earlier than reported in the modern observational era': to back up a statement like this, I'd again like to see that the model is simulating instrumentally observed climate during the relevant temporal overlap in the historical run (e.g., show Mediterranean precip./PDSI time series in model and instrumental data) before claiming that any drying has happened earlier than reported.

We will modify the sentence to "the drying trend is already observed from the pre-industrial period." and deleting "have started earlier than reported in the modern observational era".

References

Brönnimann, S.: Impact of El Niño–southern oscillation on European climate, Reviews of Geophysics, 45, https://doi.org/10.1029/2006RG000199, 2007.

Brönnimann, S., Xoplaki, E., Casty, C., Pauling, A., and Luterbacher, J.: ENSO influence on Europe during the last centuries, Climate Dynamics, 28, 181–197, https://doi.org/10.1007/s00382-006-0175-z, 2007.

Champion, A. J., Hodges, K. I., Bengtsson, L. O., Keenlyside, N. S., & Esch, M. Impact of increasing resolution and a warmer climate on extreme weather from Northern Hemisphere extratropical cyclones. Tellus A: Dynamic meteorology and oceanography, 63(5), 893-906, 2011.

Coats, S., Smerdon, J. E., Seager, R., Cook, B. I., and González-Rouco, J. F. : Megadroughts in southwestern North America in ECHO-G millennial simulations and their comparison to proxy drought reconstructions. *Journal of climate*, *26*(19), 7635-7649, 2013

Compo, G.P., J.S. Whitaker, P.D. Sardeshmukh, N. Matsui, R.J. Allan, X. Yin, B.E. Gleason, R.S. Vose, G. Rutledge, P. Bessemoulin, S. Brönnimann, M. Brunet, R.I. Crouthamel, A.N. Grant, P.Y. Groisman, P.D. Jones, M. Kruk, A.C. Kruger, G.J. Marshall, M. Maugeri, H.Y. Mok, Ø. Nordli, T.F. Ross, R.M. Trigo, X.L. Wang, S.D. Woodruff, and S.J. Worley, 2011: The Twentieth Century Reanalysis Project. Quarterly J. Roy. Meteorol. Soc., 137, 1-28. DOI: 10.1002/qj.776, 2011.

García-Herrera, R., Garrido-Perez, J.M., Barriopedro, D., Ordóñez, C., Vicente-Serrano, S.M., Nieto, R., Gimeno, L., Sorí, R. and Yiou, P.: The European 2016/17 Drought. *Journal of Climate*, *32*(11), pp.3169-3187, 2019.

Huang, B., Peter W. Thorne, et. al, 2017: Extended Reconstructed Sea Surface Temperature version 5 (ERSSTv5), Upgrades, validations, and intercomparisons. J. Climate, doi: 10.1175/JCLI-D-16-0836.1

Lehner, F., Joos, F., Raible, C. C., Mignot, J., Born, A., Keller, K. M., and Stocker, T. F. : Climate and carbon cycle dynamics in a CESM simulation from 850 to 2100 CE, Earth System Dynamics, 6, 411–434, DOI:10.5194/esd-6-411-2015, 2015.

Mariotti, A., Zeng, N., and Lau, K.-M.: Euro-Mediterranean rainfall and ENSO—a seasonally varying relationship, Geophysical research letters, 29, 59–1, https://doi.org/10.1029/2001GL014248, 2002.

Mariotti, A., Zeng, N., Yoon, J.-H., Artale, V., Navarra, A., Alpert, P., and Li, L. Z.: Mediterranean water cycle changes: transition to drier 21st cen

tury conditions in observations and CMIP3 simulations, Environmental Research Letters, 3, 044 001, https://doi.org/10.1088/1748-9326/3/4/044001, 2008.

Martín, M. L., Luna, M. Y., Morata, A., & Valero, F.: North Atlantic teleconnection patterns of low-frequency variability and their links with springtime precipitation in the western Mediterranean. International Journal of Climatology: A Journal of the Royal Meteorological Society, 24(2), 213-230, 2004.

Mukherjee, S., Mishra, A., and Trenberth, K. E.: Climate Change and Drought: a Perspective on Drought Indices, Current Climate Change 485 Reports, 4, 145–163, https://doi.org/10.1007/s40641-018-0098-x, 2018.

Naumann, G., Alfieri, L., Wyser, K., Mentaschi, L., Betts, R.A., Carrao, H., Spinoni, J., Vogt, J. and Feyen, L.: Global changes in drought conditions under different levels of warming. Geophysical Research Letters, 45(7), pp.3285-3296, 2018.

Parsons, L. A., Coats, S., & Overpeck, J. T.: The continuum of drought in Southwestern North America. Journal of Climate, 31(20), 8627-8643, 2018.

Spinoni, J., Naumann, G., & Vogt, J. V.: Pan-European seasonal trends and recent changes of drought frequency and severity. Global and Planetary Change, 148, 113-130, 2017.

Watterson, I. G.: The intensity of precipitation during extratropical cyclones in global warming simulations: a link to cyclone intensity?. Tellus A: Dynamic Meteorology and Oceanography, 58(1), 82-97, 2006.

Willmott, C. J. and K. Matsuura: Terrestrial Air Temperature and Precipitation: Monthly and Annual Time Series (1950 - 1999), http://climate.geog.udel.edu/~climate/html_pages/README.ghcn_ts2.html, 2011.

Xoplaki, E., González-Rouco, J. F., Luterbacher, J., and Wanner, H.: Mediterranean summer air temperature variability and its connection to the large-scale atmospheric circulation and SSTs, Climate Dynamics, 20, 723–739, https://doi.org/10.1007/s00382-003-0304-x, 2003.